

*Specfor*

PAGE: 1

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/918,288

DATE: 06/09/2000  
TIME: 05:01:51

INPUT SET: S35603.raw

#18  
J.D.T.

This Raw Listing contains the General  
Information Section and up to the first 5 pages.

1 SEQUENCE LISTING  
2  
3 (1) General Information

ENTERED

4  
5 (i) APPLICANT: BOIME, Irving  
6 MOYLE, William R.

7  
8 (ii) TITLE OF THE INVENTION: SINGLE-CHAIN FORMS OF THE  
9 GLYCOPROTEIN HORMONE QUARTET

10  
11 (iii) NUMBER OF SEQUENCES: 83

12  
13 (iv) CORRESPONDENCE ADDRESS:  
14 (A) ADDRESSEE: MORRISON & FOERSTER  
15 (B) STREET: 2000 Pennsylvania Avenue, NW, suite 5500  
16 (C) CITY: Washington  
17 (D) STATE: DC  
18 (E) COUNTRY: USA  
19 (F) ZIP: 20006-1888

20  
21 (v) COMPUTER READABLE FORM:  
22 (A) MEDIUM TYPE: Diskette  
23 (B) COMPUTER: IBM Compatible  
24 (C) OPERATING SYSTEM: DOS  
25 (D) SOFTWARE: FastSEQ for Windows Version 2.0

26  
27 (vi) CURRENT APPLICATION DATA:  
28 (A) APPLICATION NUMBER: 08/918,288  
29 (B) FILING DATE:  
30 (C) CLASSIFICATION:

31  
32 (vii) PRIOR APPLICATION DATA:  
33 (A) APPLICATION NUMBER: 09/282,357  
34 (B) FILING DATE:

35  
36 (A) APPLICATION NUMBER: 08/853,524  
37 (B) FILING DATE: 09-MAY-1997

38  
39 (A) APPLICATION NUMBER: 08/199,382  
40 (B) FILING DATE: 18-FEB-1994

41  
42 (viii) ATTORNEY/AGENT INFORMATION:  
43 (A) NAME: Murashige, Kate H  
44 (B) REGISTRATION NUMBER: 29,959  
45 (C) REFERENCE/DOCKET NUMBER: 29500-20050.25

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/918,288DATE: 06/09/2000  
TIME: 05:01:52

INPUT SET: S35603.raw

47  
48 (ix) TELECOMMUNICATION INFORMATION:  
49 (A) TELEPHONE: 202-887-1500  
50 (B) TELEFAX: 202-887-0763  
51 (C) TELEX:  
52  
53 (2) INFORMATION FOR SEQ ID NO:1:  
54  
55 (i) SEQUENCE CHARACTERISTICS:  
56 (A) LENGTH: 28 amino acids  
57 (B) TYPE: amino acid  
58 (C) STRANDEDNESS: single  
59 (D) TOPOLOGY: linear  
60  
61  
62 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:  
63  
64 Ser Ser Ser Ser Lys Ala Pro Pro Pro Ser Leu Pro Ser Pro Ser Arg  
65 1 5 10 15  
66 Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu Pro Gln  
67 20 25  
68  
69 (2) INFORMATION FOR SEQ ID NO:2:  
70  
71 (i) SEQUENCE CHARACTERISTICS:  
72 (A) LENGTH: 836 base pairs  
73 (B) TYPE: nucleic acid  
74 (C) STRANDEDNESS: double  
75 (D) TOPOLOGY: linear  
76  
77 (ix) FEATURE:  
78  
79 (A) NAME/KEY: Coding Sequence  
80 (B) LOCATION: 33...827  
81 (D) OTHER INFORMATION:  
82  
83 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:  
84  
85 ATGAAATCGA CGGAATCAGA CTCGAGCAA GG ATG GAG ATG TTC CAG GGG CTG 53  
86 Met Glu Met Phe Gln Gly Leu  
87 1 5  
88  
89 CTG CTG TTG CTG CTG AGC ATG GGC GGG ACA TGG GCA TCC AAG GAG 101  
90 Leu Leu Leu Leu Leu Ser Met Gly Gly Thr Trp Ala Ser Lys Glu  
91 10 15 20  
92  
93 CCG CTT CGG CCA CGG TGC CGC CCC ATC AAT GCC ACC CTG GCT GTG GAG 149  
94 Pro Leu Arg Pro Arg Cys Arg Pro Ile Asn Ala Thr Leu Ala Val Glu  
95 25 30 35  
96  
97 AAG GAG GGC TGC CCC GTG TGC ATC ACC GTC AAC ACC ACC ATC TGT GCC 197  
98 Lys Glu Gly Cys Pro Val Cys Ile Thr Val Asn Thr Ile Cys Ala  
99

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/918,288DATE: 06/09/2000  
TIME: 05:01:53

INPUT SET: S35603.raw

100	40	45	50	55	
101					
102	GGC TAC TGC CCC ACC ATG ACC CGC GTG CTG CAG GGG GTC CTG CCG GCC				245
103	Gly Tyr Cys Pro Thr Met Thr Arg Val Leu Gln Gly Val Leu Pro Ala				
104	60	65		70	
105					
106	CTG CCT CAG GTG GTG TGC AAC TAC CGC GAT GTG CGC TTC GAG TCC ATC				293
107	Leu Pro Gln Val Val Cys Asn Tyr Arg Asp Val Arg Phe Glu Ser Ile				
108	75	80		85	
109					
110	CGG CTC CCT GGC TGC CCG CGC GGC GTG AAC CCC GTG GTC TCC TAC GCC				341
111	Arg Leu Pro Gly Cys Pro Arg Gly Val Asn Pro Val Val Ser Tyr Ala				
112	90	95		100	
113					
114	GTG GCT CTC AGC TGT CAA TGT GCA CTC TGC CGC CGC AGC ACC ACT GAC				389
115	Val Ala Leu Ser Cys Gln Cys Ala Leu Cys Arg Arg Ser Thr Thr Asp				
116	105	110		115	
117					
118	TGC GGG GGT CCC AAG GAC CAC CCC TTG ACC TGT GAT GAC CCC CGC TTC				437
119	Cys Gly Gly Pro Lys Asp His Pro Leu Thr Cys Asp Asp Pro Arg Phe				
120	120	125		130	135
121					
122	CAG GAC TCC TCT TCC TCA AAG GCC CCT CCC CCC AGC CTT CCA AGC CCA				485
123	Gln Asp Ser Ser Ser Lys Ala Pro Pro Pro Ser Leu Pro Ser Pro				
124	140	145		150	
125					
126	TCC CGA CTC CCG GGG CCC TCG GAC ACC CCG ATC CTC CCC CAA GGA TCC				533
127	Ser Arg Leu Pro Gly Pro Ser Asp Thr Pro Ile Leu Pro Gln Gly Ser				
128	155	160		165	
129					
130	GGT AGC GGA TCT GGT AGC GCT CCT GAT GTG CAG GAT TGC CCA GAA TGC				581
131	Gly Ser Gly Ser Gly Ser Ala Pro Asp Val Gln Asp Cys Pro Glu Cys				
132	170	175		180	
133					
134	ACG CTA CAG GAA AAC CCA TTC TTC TCC CAG CCG GGT GCC CCA ATA CTT				629
135	Thr Leu Gln Glu Asn Pro Phe Phe Ser Gln Pro Gly Ala Pro Ile Leu				
136	185	190		195	
137					
138	CAG TGC ATG GGC TGC TTC TCT AGA GCA TAT CCC ACT CCA CTA AGG				677
139	Gln Cys Met Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro Leu Arg				
140	200	205		210	215
141					
142	TCC AAG AAG ACG ATG TTG GTC CAA AAG AAC GTC ACC TCA GAG TCC ACT				725
143	Ser Lys Lys Thr Met Leu Val Gln Lys Asn Val Thr Ser Glu Ser Thr				
144	220	225		230	
145					
146	TGC TGT GTA GCT AAA TCA TAT AAC AGG GTC ACA GTA ATG GGG GGT TTC				773
147	Cys Cys Val Ala Lys Ser Tyr Asn Arg Val Thr Val Met Gly Gly Phe				
148	235	240		245	
149					
150	AAA GTG GAG AAC CAC ACG GCG TGC CAC TGC AGT ACT TGT TAT TAT CAC				821
151	Lys Val Glu Asn His Thr Ala Cys His Cys Ser Thr Cys Tyr Tyr His				
152	250	255		260	

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/918,288DATE: 06/09/2000  
TIME: 05:01:54

INPUT SET: S35603.raw

153  
 154 AAA TCT TAAGGTACC 836  
 155 Lys Ser  
 156 265  
 157  
 158  
 159 (2) INFORMATION FOR SEQ ID NO:3:  
 160  
 161 (i) SEQUENCE CHARACTERISTICS:  
 162 (A) LENGTH: 265 amino acids  
 163 (B) TYPE: amino acid  
 164 (C) STRANDEDNESS: single  
 165 (D) TOPOLOGY: linear  
 166  
 167 (ii) MOLECULE TYPE: protein  
 168 (v) FRAGMENT TYPE: internal  
 169  
 170 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:  
 171  
 172 Met Glu Met Phe Gln Gly Leu Leu Leu Leu Leu Leu Ser Met Gly  
 173 1 5 10 15  
 174 Gly Thr Trp Ala Ser Lys Glu Pro Leu Arg Pro Arg Cys Arg Pro Ile  
 175 20 25 30  
 176 Asn Ala Thr Leu Ala Val Glu Lys Glu Gly Cys Pro Val Cys Ile Thr  
 177 35 40 45  
 178 Val Asn Thr Thr Ile Cys Ala Gly Tyr Cys Pro Thr Met Thr Arg Val  
 179 50 55 60  
 180 Leu Gln Gly Val Leu Pro Ala Leu Pro Gln Val Val Cys Asn Tyr Arg  
 181 65 70 75 80  
 182 Asp Val Arg Phe Glu Ser Ile Arg Leu Pro Gly Cys Pro Arg Gly Val  
 183 85 90 95  
 184 Asn Pro Val Val Ser Tyr Ala Val Ala Leu Ser Cys Gln Cys Ala Leu  
 185 100 105 110  
 186 Cys Arg Arg Ser Thr Thr Asp Cys Gly Gly Pro Lys Asp His Pro Leu  
 187 115 120 125  
 188 Thr Cys Asp Asp Pro Arg Phe Gln Asp Ser Ser Ser Lys Ala Pro  
 189 130 135 140  
 190 Pro Pro Ser Leu Pro Ser Pro Ser Arg Leu Pro Gly Pro Ser Asp Thr  
 191 145 150 155 160  
 192 Pro Ile Leu Pro Gln Gly Ser Gly Ser Gly Ser Ala Pro Asp  
 193 165 170 175  
 194 Val Gln Asp Cys Pro Glu Cys Thr Leu Gln Glu Asn Pro Phe Phe Ser  
 195 180 185 190  
 196 Gln Pro Gly Ala Pro Ile Leu Gln Cys Met Gly Cys Cys Phe Ser Arg  
 197 195 200 205  
 198 Ala Tyr Pro Thr Pro Leu Arg Ser Lys Lys Thr Met Leu Val Gln Lys  
 199 210 215 220  
 200 Asn Val Thr Ser Glu Ser Thr Cys Cys Val Ala Lys Ser Tyr Asn Arg  
 201 225 230 235 240  
 202 Val Thr Val Met Gly Gly Phe Lys Val Glu Asn His Thr Ala Cys His  
 203 245 250 255  
 204 Cys Ser Thr Cys Tyr Tyr His Lys Ser  
 205 260 265

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/918,288DATE: 06/09/2000  
TIME: 05:01:55

INPUT SET: S35603.raw

206  
207 (2) INFORMATION FOR SEQ ID NO:4:  
208209 (i) SEQUENCE CHARACTERISTICS:  
210 (A) LENGTH: 834 base pairs  
211 (B) TYPE: nucleic acid  
212 (C) STRANDEDNESS: double  
213 (D) TOPOLOGY: linear214  
215 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:  
216217  
218 TCCGGATTAG CTTGAGATGG ATCCGGTACC TTAAGATTTG TGATAATAAC AAGTACTGCA 60  
219 GTGGCACGCC GTGTGGTTCT CCACCTTGAA ACCCCCCATT ACTGTGACCC TGTTATATGA 120  
220 TTTAGCTACA CAGCAAGTGG ACTCTGAGGT GACGTTCTT TGGACCAACA TCGTCTTCTT 180  
221 GGACCTTAGT GGAGTGGGAT ATGCTCTAGA GAAGCAGCAG CCCATGCACT GAAGTATTGG 240  
222 GGCACCCGGC TGGGAGAAGA ATGGGTTTTC CTGTAGCGTG CATTCTGGGC AATCCTGCAC 300  
223 ATCAGGAGCG CTACCAGATC CGCTACCGGA TCCTTGGGG AGGATGGGG TGTCCGAGGG 360  
224 CCCCGGGAGT CGGGATGGGC TTGGAAGGCT GGGGGGAGGG GCCTTGAGG AAGAGGAGTC 420  
225 CTGGAAGCGG GGGTCATCAC AGGTCAAGGG GTGGTCCTTG GGACCCCCGC AGTCAGTGGT 480  
226 GCTGCGGCAG CAGAGTGCAC ATTGACAGCT GAGAGCCACG GCGTAGGAGA CCACGGGGTT 540  
227 CACGCCGCGC GGGCAGCCAG GGAGCCGGAT GGACTCGAAG CGCACATCGC GGTAGTTGCA 600  
228 CACCACCTGA GGCAGGGCCG GCAGGACCCC CTGCAGCACG CGGGTCATGG TGGGGCAGTA 660  
229 GCCGGCACAG ATGGTGGGTG TGACGGTGAT GCACACGGGG CAGCCCTCCT TCTCCACAGC 720  
230 CAGGGTGGCA TTGATGGGGC GGACCCGTGG CCGAAGCGGC TCCTTGGATG CCCATGTCCC 780  
231 GCCCATGCTC AGCAGCAGCA ACAGCAGCAG CCCCTGGAAC ATCTCCATCC TTGG 834  
232233 (2) INFORMATION FOR SEQ ID NO:5:  
234235 (i) SEQUENCE CHARACTERISTICS:  
236 (A) LENGTH: 743 base pairs  
237 (B) TYPE: nucleic acid  
238 (C) STRANDEDNESS: double  
239 (D) TOPOLOGY: linear240 (ix) FEATURE:  
241242 (A) NAME/KEY: Coding Sequence  
243 (B) LOCATION: 33...734  
244 (D) OTHER INFORMATION:  
245246 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:  
247248  
249 ATGAAATCGA CGGAATCAGA CTCGAGCAA GG ATG GAG ATG TTC CAG GGG CTG 53  
250 Met Glu Met Phe Gln Gly Leu  
251 1 5252  
253 CTG CTG TTG CTG CTG AGC ATG GGC GGG ACA TGG GCA TCC AAG GAG 101  
254 Leu Leu Leu Leu Leu Ser Met Gly Gly Thr Trp Ala Ser Lys Glu  
255 10 15 20256  
257 CCG CTT CGG CCA CGG TGC CGC CCC ATC AAT GCC ACC CTG GCT GTG GAG 149  
258 Pro Leu Arg Pro Arg Cys Arg Pro Ile Asn Ala Thr Leu Ala Val Glu

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
PATENT APPLICATION **US/08/918,288**

DATE: 06/09/2000  
TIME: 05:01:57

***INPUT SET: S35603.raw***

Line

Error

Original Text